

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 10-200858

(43)Date of publication of application : 31.07.1998

(51)Int.Cl.

H04N 5/92
G11B 20/10
G11B 20/12
G11B 20/12
H04N 5/928
H04N 7/24

(21)Application number : 09-001217

(71)Applicant : SONY CORP

(22)Date of filing : 08.01.1997

(72)Inventor : SHIGENOBU MASAHIRO
INOUE KOJI

(54) ENCODER, ENCODING METHOD, DECODER, DECODING METHOD AND RECORDING MEDIUM FOR VIDEO SIGNAL AND SOUND SIGNAL

(57)Abstract:

PROBLEM TO BE SOLVED: To synchronize and record video signals and sound signals by a GOP (Group of Pictures) unit by digitizing and encoding the sound signals to include a fixed sample number in the GOP unit of compression video data at the time of encoding and compressing the video signals by MPEG and encoding and compressing the sound signals by ATRAC.

SOLUTION: The sampling frequency of the sound signals is set to the integrally multiple of a GOP frequency (2000/1001Hz). Actually, generally used 48.0kHz is turned to the sampling frequency of the sound signals. The sample number of sound data in 1 GOP obtained from the sampling frequency is 24, 024 pieces, and since a fraction is generated when the sample number of the sound data are sectioned by 512 pieces, the valid sample number is actually sectioned by 511 pieces. Further, one is added to the 7 pieces of the samples, the sample number of 512 pieces is attained and the sound data are almost equally arranged corresponding to 1 GOP and synchronized as the whole.

(19)日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平10-200858

(43)公開日 平成10年(1998)7月31日

(51)Int.Cl. ⁸	識別記号	F I	
H 0 4 N 5/92		H 0 4 N 5/92	H
G 1 1 B 20/10	3 0 1	G 1 1 B 20/10	3 0 1 Z
	1 0 2	20/12	1 0 2
	1 0 3		1 0 3
H 0 4 N 5/928		H 0 4 N 5/92	E

審査請求 未請求 請求項の数 8 OL (全 11 頁) 最終頁に続く

(21)出願番号 特願平9-1217

(22)出願日 平成9年(1997)1月8日

(71)出願人 000002185

ソニー株式会社

東京都品川区北品川6丁目7番35号

(72)発明者 重信 正大

東京都品川区北品川6丁目7番35号 ソニー株式会社内

(72)発明者 井上 光司

東京都品川区北品川6丁目7番35号 ソニー株式会社内

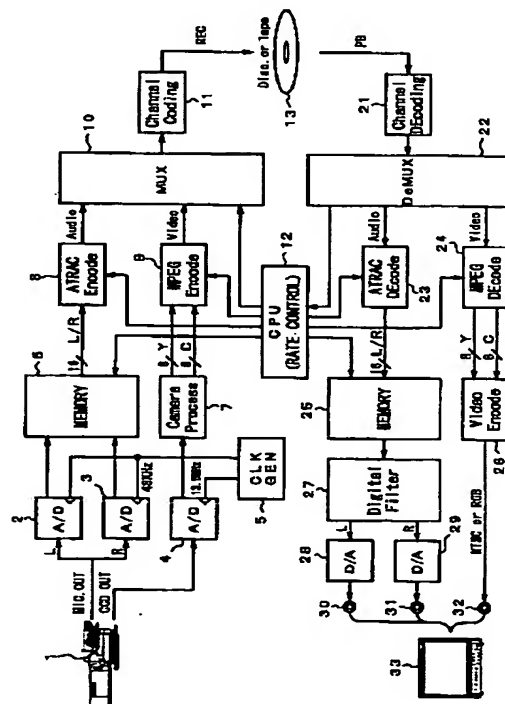
(74)代理人 弁理士 杉浦 正知

(54)【発明の名称】 映像信号および音声信号の符号化装置、符号化方法、復号装置および復号方法並びに記録媒体

(57)【要約】

【課題】 MPEG2で圧縮された映像データと、ATRACで符号化された音声データとの同期を取り記録媒体に記録／再生を行うとができる。

【解決手段】 記録媒体13には、GOP単位で記録／再生を行う。GOPは、ヘッダ情報、映像データおよび音声データから構成される。1GOPは、例えば15フレーム(約0.5秒)からなり、音声データのサンプル数を整数個とするために、サンプリング周波数を48.0kHzと設定される。このときのサンプル数が24,024個であり、このサンプル数を47個のサウンドグループに分ける。1サウンドグループのサンプル数は、511個または512個となる。511個のサンプル数のサウンドグループを扱うときは、例外処理として、1つ前のサンプルをダミーとして用い、512個のサウンドグループとして処理を行う。このサウンドグループが511個か512個かは、ヘッダ情報に記録される。



【特許請求の範囲】

【請求項1】 映像データをMPEGで符号化し、圧縮映像データを生成する映像符号化手段と、

上記圧縮映像データのGOP単位で一定となるサンプル数が含まれるように、上記音声信号のサンプリング周波数を設定し、上記サンプリング周波数で上記音声信号をデジタル化するA/D変換手段と、

上記A/D変換手段からの音声データを符号化し、符号化音声データを生成する音声符号化手段と、

上記符号化音声データを複数のグループに分け、上記圧縮映像データおよび上記複数のグループに分けられた上記符号化音声データを上記GOP毎に記録媒体に記録する記録手段とからなることを特徴とする映像信号および音声信号の符号化装置。

【請求項2】 請求項1において、上記グループ単位で映像データと同期して音声データをアフターレコーディングが行えるようにしたことを特徴とする映像信号および音声信号の符号化装置。

【請求項3】 請求項1において、上記グループのサンプル単位で映像データと同期して音声データをアフターレコーディングが行えるようにしたことを特徴とする映像信号および音声信号の符号化装置。

【請求項4】 記録媒体に記録された圧縮映像データおよび符号化音声データをGOP単位で読み出す読み出し手段と、

上記圧縮映像データをMPEGで復号し、映像データを復号する映像復号手段と、

上記符号化音声データを上記GOPのヘッダ情報に基づいて復号し、音声データを復号する音声復号手段と、

上記映像データおよび上記音声データを上記GOP単位で同期させる手段とからなることを特徴とする映像信号および音声信号の復号装置。

【請求項5】 請求項4において、上記音声復号手段は、上記ヘッダ情報から上記符号化音声データを構成するグループのサンプル数を読み取り、上記読み取られたサンプル数に基づいて上記符号化音声データを復号するようにしたことを特徴とする映像信号および音声信号の復号装置。

【請求項6】 映像データをMPEGで符号化し、圧縮映像データを生成するステップと、

上記圧縮映像データのGOP単位で一定となるサンプル数が含まれるように、上記音声信号のサンプリング周波数を設定し、上記サンプリング周波数で上記音声信号をデジタル化するステップと、

上記デジタル化された音声データを符号化し、符号化音声データを生成するステップと、

上記符号化音声データを複数のグループに分け、上記圧縮映像データおよび上記複数のグループに分けられた

上記符号化音声データを上記GOP毎に記録媒体に記録するステップとからなることを特徴とする映像信号および音声信号の符号化方法。

【請求項7】 記録媒体に記録された圧縮映像データおよび符号化音声データをGOP毎に読み出すステップと、

上記圧縮映像データをMPEGで復号し、映像データを復号するステップと、

上記符号化音声データを上記GOPのヘッダ情報に基づいて復号し、音声データを復号するステップと、

上記映像データおよび上記音声データをGOP単位で同期させるステップとからなることを特徴とする映像信号および音声信号の復号方法。

【請求項8】 MPEGが施された圧縮映像データと、符号化が施された符号化音声データとを記録するようにした記録媒体において、

上記圧縮映像データおよび上記符号化音声データを記録する記録部を有し、

上記記録部には、GOPを単位として、ヘッダ情報と、上記圧縮映像データと、上記符号化音声データが記憶され、

上記符号化音声データは、上記GOP単位で一定となるようにデジタル化された音声データを符号化すると共に、上記GOP単位のサンプル数が複数のグループに分けられるように形成されたことを特徴とする記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】この発明は、異なる単位で圧縮が施された映像信号および音声信号を同期させることが可能な映像信号および音声信号の符号化装置、符号化方法、復号装置および復号方法並びに記録媒体に関する。

【0002】

【従来の技術】現在、音声信号に圧縮処理を施し、ディスクまたはテープ等の記録媒体に記録／再生を行う手法として、標準化されたATRAC (Adaptive Transform Acoustic Coding) が知られている。このATRACは、MD (Mini Disk) に採用されている。また、映像信号と音声信号とに圧縮処理を施し、ディスクまたはテープ等の記録媒体に記録／再生を行う手法として、標準化されたMPEG2 (Moving Picture Experts Group phase 2) が知られている。このMPEG2は、DVD (Digital Video Disk) やデジタル放送等に採用されている。

【0003】

【発明が解決しようとする課題】映像信号を圧縮してディスクやテープ等のメディアに記録する方式が多数提案されている。その1つが上述したMPEG2であり、このMPEG2では、映像信号の圧縮が15フレーム（約

0.5秒)単位で行われるものが主流となりつつある。この圧縮の単位をGOP(Group of Pictures)と呼ぶ。NTSC方式において、GOPの時間は、正確には、 $1001/2000$ 秒と規定されており、この単位で映像信号と音声信号の同期を取るためには、音声信号のサンプリング周波数とGOP周波数($2000/1001$ Hz)がある関係、例えば整数比の関係になければいけない。

【0004】具体的には、MDやCD等で用いられるサンプリング周波数は、 44.1 kHzであるが、この周波数では、1GOP($1001/2000$ 秒)の中のサンプル数が22,072.05個と小数になり、映像信号と音声信号の同期を取るのは困難であった。すなわち、これら標準化された圧縮の手法において、相互間で利用することは困難であった。

【0005】従って、この発明の目的は、例えば上述のATracおよびMPEG2のように、異なる単位で圧縮された映像信号および音声信号を同期させることが可能な映像信号および音声信号の符号化装置、符号化方法、復号装置および復号方法並びに記録媒体を提供することにある。

【0006】

【課題を解決するための手段】請求項1に記載の発明は、映像データをMPEGで符号化し、圧縮映像データを生成する映像符号化手段と、圧縮映像データのGOP単位で一定となるサンプル数が含まれるように、音声信号のサンプリング周波数を設定し、サンプリング周波数で音声信号をデジタル化するA/D変換手段と、A/D変換手段からの音声データを符号化し、符号化音声データを生成する音声符号化手段と、符号化音声データを複数のグループに分け、圧縮映像データおよび複数のグループに分けられた符号化音声データをGOP毎に記録媒体に記録する記録手段とからなることを特徴とする映像信号および音声信号の符号化装置である。

【0007】請求項4に記載の発明は、記録媒体に記録された圧縮映像データおよび符号化音声データをGOP単位で読み出す読み出し手段と、圧縮映像データをMPEGで復号し、映像データを復号する映像復号手段と、符号化音声データをGOPのヘッダ情報に基づいて復号し、音声データを復号する音声復号手段と、映像データおよび音声データをGOP単位で同期させる手段とからなることを特徴とする映像信号および音声信号の復号装置である。

【0008】請求項6に記載の発明は、映像データをMPEGで符号化し、圧縮映像データを生成するステップと、圧縮映像データのGOP単位で一定となるサンプル数が含まれるように、音声信号のサンプリング周波数を設定し、サンプリング周波数で音声信号をデジタル化するステップと、デジタル化された音声データを符号化し、符号化音声データを生成するステップと、符号化

音声データを複数のグループに分け、圧縮映像データおよび複数のグループに分けられた符号化音声データをGOP毎に記録媒体に記録するステップとからなることを特徴とする映像信号および音声信号の符号化方法である。

【0009】請求項7に記載の発明は、記録媒体に記録された圧縮映像データおよび符号化音声データをGOP毎に読み出すステップと、圧縮映像データをMPEGで復号し、映像データを復号するステップと、符号化音声データをGOPのヘッダ情報に基づいて復号し、音声データを復号するステップと、映像データおよび音声データをGOP単位で同期させるステップとからなることを特徴とする映像信号および音声信号の復号方法である。

【0010】請求項8に記載の発明は、MPEGが施された圧縮映像データと、符号化が施された符号化音声データとを記録するようにした記録媒体において、圧縮映像データおよび符号化音声データを記録する記録部を有し、記録部には、GOPを単位として、ヘッダ情報と、圧縮映像データと、符号化音声データが記憶され、符号化音声データは、GOP単位で一定となるようにデジタル化された音声データを符号化すると共に、GOP単位のサンプル数が複数のグループに分けられるように形成されたことを特徴とする記録媒体である。

【0011】この発明は、映像信号をMPEGで圧縮を施し、音声信号をATracで圧縮を施した後、1GOPを単位として、同期が取られ、記録媒体に記録することができる。また、記録媒体に記録されたデータは、1GOP毎に再生することができ、そのときの映像信号と音声信号との同期も容易にとることができる。そして、記録媒体に記録されたデータは、1GOP毎にアフレコ(アフターレコーディング)を行うこともでき、映像信号および音声信号、また音声信号のみでも容易にアフレコを行うことができる。

【0012】

【発明の実施の形態】以下、この発明の一実施例を図面を参照して説明する。この一実施例では、ATracにより圧縮されるオーディオデータのサンプリング周波数をDVD等で用いられている48.0 kHzとする。48.0 kHzのサンプリング周波数では、1GOPの中に24,024個の整数個のサンプリング点が存在することになる。それによって、映像信号と音声信号との同期をとることを可能とするものである。なお、サンプリング周波数としては、48.0 kHzに限らず他の周波数、例えば96.0 kHzを使用しても良い。

【0013】この発明の一実施例の全体ブロック図を図1に示す。まず、記録側において、1で示すビデオカメラから音声信号および映像信号が出力される。ビデオカメラ1から出力される音声信号はMIC. OUTとして、Lチャンネルの音声信号がA/D変換器2へ供給され、Rチャンネルの音声信号がA/D変換器3へ供給さ

れる。A/D変換器2および3は、クロック発生器（CLK GEN）5から供給された48.0 kHzの周波数のサンプリングクロックによって、デジタル化を行う。

【0014】A/D変換器2および3からの音声データは、メモリ6へ供給される。メモリ6は、映像データと音声データのタイミングを調整するためのものである。CPU12の制御によって、メモリ6から音声データがATRACエンコーダ8へ供給される。音声データは、例えば16ビットからなるリニアPCM（Pulse Code Modulation）データであり、時分割されたLチャンネルとRチャンネルの音声信号である。ATRACエンコーダ8では、CPU12からのコントロール信号に基づいて、後述するようにGOPに対応する音声データが47個のサウンドグループに分けられる。サウンドグループに分けられた音声信号は、マルチプレクサ（MUX）10へ供給される。

【0015】映像信号は、ビデオカメラ1からCCD OUTとして、A/D変換器4へ供給される。A/D変換器4には、クロック発生器5から13.5 kHzの周波数のサンプリングクロックが供給され、映像信号がデジタル化される。デジタル化された映像データは、カメラプロセス回路7へ供給される。カメラプロセス回路7では、 γ 補正およびアバーチャ補正等のカメラ信号処理が行われ、輝度データYおよび色差データCが生成される。この色差データCは、時分割された色差データCb（青の色差データ）および色差データCr（赤の色差データ）からなる。輝度データYおよび色差データCは、それぞれ8ビットのデータとしてMPEGエンコーダ9へ供給される。MPEGエンコーダ9では、CPU12からのGOPを指示する信号等のコントロール信号に基づいて、輝度データYおよび色差データCがMPEG2で符号化される。符号化された映像データは、マルチプレクサ10へ供給される。

【0016】マルチプレクサ10は、符号化された音声データと符号化された映像データとCPU12からのヘッダ情報とが供給され、ヘッダ情報の付加、音声信号の例外処理を行う。マルチプレクサ10からは、図2に示すように映像信号と音声信号とヘッダ情報が合成された記録データが出力される。記録データは、チャンネルコーディング回路11へ供給される。チャンネルコーディング回路11では、（1，7）RLLの変調、スクランブルおよびエラー訂正のエンコーディング等の処理が行われる。これらの処理が施されたデータは、ビットストリームで記録媒体13へ記録される。この記録媒体13は、磁気または光テープ、MO（光磁気）ディスク等の記録可能なディスク、メモ리카ード等の半導体メモリ等である。

【0017】次に再生側において、記録媒体13からビットストリームで読み出されたデータは、チャンネルデ

コーディング回路21へ供給される。チャンネルデコーディング回路21では、（1，7）RLLの復調、ディスクランブルおよびエラー訂正等の処理が行われる。これらの処理が行われたデータは、ディマルチプレクサ22へ供給される。ディマルチプレクサ22では、供給されたデータからヘッダ情報、映像データおよび音声データが分離され、音声データはATRACデコーダ23へ供給され、映像データはMPEGデコーダ24へ供給される。また、ディマルチプレクサ22で分離されるヘッダ情報は、CPU12へ供給される。CPU12では、供給されたヘッダ情報から音声信号の例外処理の位置等の復号に必要な情報が検出され、その情報はコントロール信号としてATRACデコーダ23およびMPEGデコーダ24へ供給される。

【0018】ATRACデコーダ23では、CPU12からのコントロール信号に基づいて、ATRACで符号化された音声データが復号される。復号された音声データは、時分割されたLチャンネルとRチャンネルの音声信号であり、メモリ25へ供給される。メモリ25は、CPU12の制御によって、映像データと音声データのタイミングを調整するものである。メモリ25から読み出された音声データは、デジタルフィルタ27へ供給される。デジタルフィルタ27は、後段のD/A変換器28および29に対するプリフィルタであり、Lチャンネルの音声データは、D/A変換器28へ供給され、Rチャンネルの音声データは、D/A変換器29へ供給される。アナログ化された音声信号は、それぞれ出力端子30および31を介してTVモニタ33の音声信号入力端子へ供給される。

【0019】ディマルチプレクサ22から映像データが供給されたMPEGデコーダ24では、CPU12からのGOPを指示する信号等のコントロール信号に基づいて、MPEG2の復号を行う。復号された輝度データYおよび色差データCは、それぞれ8ビットのデータとしてビデオエンコーダ26へ供給される。復号された色差データCは、時分割された色差データCrおよび色差データCbである。ビデオエンコーダ26では、供給された輝度データYと色差データCからNTSC信号またはRGB信号が復号される。ビデオエンコーダ26には、D/A変換器が含まれビデオエンコーダ26からアナログのNTSC信号またはRGB信号が出力される。NTSC信号またはRGB信号は、出力端子32を介してTVモニタ33の映像信号入力端子へ供給される。

【0020】ここで、この発明の要旨となる、MPEG2によって圧縮された映像データとATRACによって圧縮された音声データとを同期して記録する方法の一例を具体的に説明する。音声信号のサンプリング周波数は、GOP周波数（2000/1001Hz）の整数倍の周波数が設定される。具体的には、GOP周波数とサンプリング周波数の最小公倍数が小さく、一般的に用いら

れている48.0kHzを音声信号のサンプリング周波数とする。このサンプリング周波数から求められる1GOP中の音声データのサンプル数は、24,024個である。ATRACエンコーダは、8通りのモードが定義されており、この実施例では、一例としてコーディングモード:ATRAC(1),Level1,ChannelNo.=2について説明する。

【0021】このコーディングモードは、音声データの片チャンネルの512個のサンプル数を212バイトの圧縮データにエンコードするものである。この512個のサンプル数のATRACの処理単位をサウンドグループ(SG)と呼ぶことにする。以下、説明を容易とするためにLチャンネルのみについて説明する。24,024個の音声データのサンプル数を512個のサンプル単位に区切ると、46個のサウンドグループ+472個の音声データのサンプル数となる。従って、サウンドグループを47個分用意すれば、端数は生じるが、1GOP分の音声データを格納することができる。

【0022】この48.0kHzのサンプリング周波数でサンプリングされた1GOP分の音声データのサンプル数の24,024個と1GOP分の音声データのサンプルとを比較すると、1GOP分の音声データのサンプル

数は、 $47(SG) \times 512(\text{サンプル}) = 24064(\text{サンプル})$

であるから、1GOPのサンプル数(24,024個)に対して40個のサンプルの端数が生じる。

【0023】この40個のサンプルを吸収するために、この実施例では、47個のサウンドグループのうち40個のサウンドグループについて、実際に使用される音声データのサンプル数を1個減らして511個として、符号化が実行される。このように、有効なサンプル数を1個減少させる処理が例外処理である。残りの7個のサウンドグループについては、音声データのサンプル数を512個として、符号化が実行される。この512個の音声データのサンプル数からなるサウンドグループは、出来るだけ47個のサウンドグループ中に均等に配置させる。例えば、0、7、14、21、27、34、41番目に512個の音声データのサンプルからなるサウンドグループを配置させることによって、ほぼ均等に音声データを配置することができる。

【0024】 $40(SG) \times 511(\text{サンプル}) + 7(SG) \times 512(\text{サンプル}) = 24024(\text{サンプル})$

となり、1GOP分の音声データのサンプル数となる。すなわち、1GOP(約0.5秒)を単位として、映像データおよび音声データが完結することになる。

【0025】ここで、音声データのサンプル数が511個のサウンドグループをATRACでエンコードするときの手法を簡単に説明する。上述したように、この実施

例で使われているATRACのエンコードでは、1サウンドグループについて、512個の音声データのサンプル数が必要である。しかしながら、1GOP分の音声データのサンプル数が24,024個のため、サンプル数が511個のサウンドグループを40個作るようにした。このサンプル数が511個のサウンドグループをATRACでエンコードする場合、1つ前のサウンドグループの512番目のサンプルをダミーデータとして使用する。そのダミーデータ1個を加えて、サンプル数が512個となったサウンドグループがATRACでエンコードされる。

【0026】次に再生について、説明する。再生時、すなわちデコード時は、ATRACのデコーダから出力される512個のサンプルのうち、後半511個のサンプルを実データとして使用すれば良い。最初の1サンプルは、エンコード時、ダミーデータという扱いだったので捨てる。サウンドグループの境目は、後述するように、オーバーラップ分があり、窓関数の掛ける処理がなされるので、波形の連続性は、保たれる。よって、ダミーデータとなる1サンプルの処理は、従来に比べてオーバーラップのサンプル数が1サンプルだけ増えると考えれば良い。ただし、どのサウンドグループが511個のサンプルを出力するのか、または512個のサンプルを出力するのかは、エンコード時にヘッダ情報に記録しておく必要がある。

【0027】図2に示すように、記録媒体には、上述したようにGOPのヘッダ情報、15フレームからなる1GOP分(約0.5秒)の映像データおよび約0.5秒(1GOP分)に相当する47サウンドグループからなる音声データが最小単位として、扱われる。このとき、1サウンドグループは、511個または512個のサンプルからなり、1GOPで24,024個のサンプルが存在する。そして、15フレームからなる映像データは、MPEG2で圧縮処理が施され、47サウンドグループからなる音声データは、ATRACで圧縮処理が施される。また、例外処理によって有効な音声データのサンプル数が511個とされたサウンドグループの位置の情報がヘッダ情報に記憶される。このヘッダ情報と、圧縮が施された映像データと音声データとからなる記録データ(ビットストリーム)が(1,7)RLLの変調され、記録媒体へ記録される。

【0028】この記録媒体に対して、1GOP単位で映像データと音声データとが記録される。まず、ヘッダ情報に続いて、1GOP分の圧縮映像データ、さらに続いて2チャンネル分の音声データが1GOPの期間で記録される。ヘッダ情報には、どのサウンドグループが511個または512個のサンプル数かが記録されている。この記録される音声データは、 $(47(SG) \times 2(ch) = 94)$ サウンドグループからなる。さらに、GOP単位で映像データと音声データの位相は、常に合致し

ている。

【0029】そして、ユーザからGOP単位で再生箇所の指定を受け取ると、指定場所からのビットストリームを復調した後、復調されたビットストリームは、映像データおよび音声データのデコーダへ入力される。映像データのデコーダは、MPEGデコーダ24であり、音声データのデコーダは、ATRACデコーダ23である。ATRACデコーダ23の出力は、一旦メモリ25に入力される。この音声データは、デコードがなされた後、メモリ25に記憶され、映像データとのタイミング合わせを行って後段へ出力される。ATRACデコーダ23およびMPEGデコーダ24の出力は、GOP単位で完結するので、一度同期を取ればずれることはない。

【0030】また、ATRACデコーダ23から受け取る音声データは、サウンドグループ当り512個のサンプル数であったり、511個のサンプル数であったりするが、上述したように、この例外処理のなされたサウンドグループが1GOP毎にヘッダ情報に書き込まれているので、これを読み込むことによって、1GOP内のサンプル数が丁度24、024個である音声データが得られる。

【0031】このように、この一実施例では、1GOPで映像データおよび音声データが完結しているので、インサートやアフレコ等をGOP単位で行うことは全く問題なくできる。

【0032】図1中のATRACエンコーダ8の詳細なブロック図を図3に示す。メモリ6からの各サンプルが16ビットからなるPCMデータ（リニアPCM）が入力端子41を介してATRAC8へ供給される。このPCMデータは、リングバッファ46へ供給される。また、入力端子42からサウンドグループのサンプル数情報がATRACエンコーダ8へ供給されると共に、出力端子45から後段のマルチプレクサ10へ供給される。このサウンドグループのサンプル数情報は、アドレス・コントロール回路47へ供給される。

【0033】アドレス・コントロール回路47では、供給されたサウンドグループのサンプル数情報からこのサウンドグループが511個のサンプルからなるものか、512個のサンプルからなるものかが判断される。このアドレス・コントロール回路47でこのサウンドグループは、511個のサンプルからなるサウンドグループであると判断されると、リングバッファ46から読み出される開始アドレスが1つ前にずらされる。リングバッファ46は、図4Aに示すように、サウンドグループ3個分の容量からなり、入力PCMデータが記憶される。

【0034】乗算器48では、図4Aに示すように、リングバッファ46から読み出されたPCMデータと窓関数とが乗算される。この図4Aにおいて、実線部は、今回の1サウンドグループ分の処理であり、破線部は、次の1サウンドグループ分の処理である。その乗算結果

は、帯域分割フィルタ49へ供給され、高域成分（H）と中域成分（M）および低域成分（L）へ分割される。高域成分（H）は、MDCT（Modified Discrete Cosine Transform）51へ供給され、その他の成分は、帯域分割フィルタ50へ供給される。帯域分割フィルタ50では、供給された成分が中域成分（M）と低域成分（L）へ分割される。分割された中域成分（M）は、MDCT52へ供給され、低域成分（L）は、MDCT53へ供給される。

【0035】MDCT51、52および53では、供給された高域成分、中域成分および低域成分に対して変形DCTの演算が施され、各周波数成分に分解される。MDCT51、52および53の出力は、ビット割当計算回路54および正規化／量子化回路55へ供給される。ビット割当計算回路54では、供給された各成分から計算によって、割り当てられる量子化ビット数が算出される。量子化ビット数を表す情報は、正規化／量子化回路55へ供給されるとともに、出力端子43から出力され、後段のマルチプレクサ10へ供給される。

【0036】正規化／量子化回路55では、量子化ビット数の情報から量子化ビット数を決定し、その量子化ビット数を用いてMDCT51、52および53から供給された高域成分、中域成分および低域成分（AC成分）に対して量子化が実行される。さらに、量子化がなされた各AC成分に対して正規化が実行される。また、正規化／量子化回路55からは、処理後のAC成分がスペクトル情報として、DC成分とともに出力され、この出力が出力端子44を介して後段のマルチプレクサ10へ供給される。

【0037】ここで、音声データのサンプル数が511個のサウンドグループをATRACでエンコードする手法を具体的に説明する。この実施例で使われているATRACのエンコードは、1サウンドグループすなわち512個の音声データのサンプル数を単位として実行される。しかしながら、実際には、この512個のサンプル以外に前後をそれぞれ256個のサンプルを加えて、合計1024個のサンプルのデータを持ってきて、窓関数を掛けてATRACのエンコードは行われる。この窓関数は、図4に示すように、1024個の全サンプル数の両端を0でうめるような関数である。

【0038】当然、511個のサンプルを扱うときも1024個のサンプルをATRACエンコーダの入力として用いる。ただし、この場合の有効サンプル数は、1サンプル減り、258番目から768番目までの511個のサンプルとなる。すなわち、図4Bに示すように、n番目のサウンドグループをエンコードする場合、中心となるサウンドグループのサンプル数の512個と、そのサウンドグループの前後の256個のサンプルの合計1,024個がATRACエンコーダに入力される。

【0039】しかしながら、次のn+1番目のサウンド

グループをエンコードする場合、サウンドグループのサンプル数は511個のため、 n 番目のサウンドグループの512番目を含めてサンプルを512個とし、その512個のサウンドグループの前後の256個のサンプルの合計1,024個がATRACエンコーダに入力される。すなわち、サウンドグループのサンプルが511個の場合、1サンプルだけ重なって、処理されるが、この重なった部分は、 n 番目のサウンドグループの符号化の処理においては、本来の処理であり、今回($n+1$ 番目)のサウンドグループの符号化の処理では、ダミーデータとして処理される。

【0040】図1中のATRACデコーダ24の詳細なブロック図を図5に示す。入力端子61から量子化ビット数の情報および正規化係数の情報がATRACデコーダ24へ供給され、同様に入力端子62からスペクトル情報がATRACデコーダ24へ供給される。また、入力端子63からサウンドグループのサンプル数情報がATRACデコーダ24へ供給されると共に、出力端子65を介してCPU12へ供給される。

【0041】ATRACデコーダ24に供給された量子化ビット数の情報、正規化係数の情報およびスペクトル情報がスペクトル復元回路71へ供給される。スペクトル復元回路71では、供給されたスペクトル情報、正規化係数の情報および量子化ビット数の情報からAC成分が生成される。生成されたAC成分の高域成分(H)は、IMDCT(逆MDCT)72へ供給される。同様にAC成分の中域成分(M)は、IMDCT73へ供給される。さらに、低域成分(L)のAC成分およびDC成分は、IMDCT74へ供給される。

【0042】IMDCT72、73および74では、供給された各成分に対して逆MDCT処理が施される。IMDCT72で逆MDCT処理により得られた高域成分の信号は、帯域合成フィルタ76へ供給される。また、IMDCT73および74で逆MDCT処理により得られた中域成分および低域成分の信号は、帯域合成フィルタ75へ供給される。帯域合成フィルタ75では、中域成分の信号と低域成分の信号とが合成され、その合成信号は、帯域合成フィルタ76へ供給される。帯域合成フィルタ76では、高域成分と合成信号との合成が行われ、その合成信号は、乗算器77へ供給される。

【0043】乗算器77では、窓関数と合成信号との乗算が行われ、その乗算結果は、リングバッファ78に記録される。この窓関数は、図4に示すように、1024個の全サンプル数の両端を0でうめるような関数である。図4Aに示すように、今回の窓関数により得られた乗算結果と、次の窓関数により得られた乗算結果とのオーバーラップとをリングバッファ78に書き込む時は、前に書き込んだデータを読み出し、読み出したデータと現データとを加算する。このように合成された窓関数は、リングバッファ78に記録される。このリングバ

ッファ78は、サウンドグループ3個分の容量からなる。アドレス・コントロール回路79では、入力端子63から供給されたサウンドグループのサンプル数情報からリングバッファ78に記録されたサウンドグループが511個のサンプルからなるものか、512個のサンプルからなるものかが判断される。このアドレス・コントロール回路79でこのサウンドグループは、511個のサンプルからなるサウンドグループであると判断されると、リングバッファ78の開始アドレスが1つ前にずらされる。リングバッファ78から16ビットからなるPCMデータが読み出されるときに、ダミーデータの1サンプル分を取り除き、出力端子64から後段のメモリ25へ供給される。

【0044】この一実施例を適用して動き予測を用いた動画圧縮メディアの場合、映像データあるいは映像データおよび音声データの編集は、GOP単位で行うことが一般的なので、上述の通り問題は生じない。しかしながら、音声データのみのアフレコ等の場合、GOP単位とは限らない。この音声データのみのアフレコの場合の1つの方法は、サウンドグループ(511/512個のサンプル)単位とすることである。これにより、より細かい単位での編集作業が可能となる。ただし、サンプル数が511個か512個かを知る必要があるため、編集開始点がGOP内の何サウンドグループ目かを正しく認識する必要がある。

【0045】さらに細かい編集作業を要求される場合には、記録時に編集開始箇所および編集終了箇所が含まれるサウンドグループのデータが読み込まれ、そのデータがデコードされた後、メモリに取り込まれる。そして、更新すべきデータがメモリに書き込まれ、編集がなされた後、サウンドグループは、再度エンコードされて記録媒体へ書き込まれる。という方法がある。この方法の場合、1サンプル単位で編集が可能となるが、エンコード/デコードを繰り返すため、編集開始点および編集終了点を含むサウンドグループが編集の度に劣化する欠点を持つ。しかしながら、最初から非可逆な圧縮を行っている点、さらにサウンドグループが約10msと極めて短い点から大きな問題はない。一般的には、サウンドグループ単位のアフレコで十分である。

【0046】上述の実施例では、1サウンドグループのデータを512個のサンプルと仮定して説明したが、1024個のサンプルの時も同様の考え方を適用できる。その場合、1GOPが15フレームならば、GOP内のサウンドグループは1チャンネル24となり、ダミーデータの個数を変更する必要がある。さらに、1サウンドグループあたりのサンプル数が別の値をとるとき、1GOPのフレーム数が別の値をとるとき、音声データのサンプリング周波数が異なる時も1GOPあたりの音声データのサンプル数が整数になりさえすれば、同様の考え方を適用できる。このときも、GOP内のサウンドグル

ープ数、サウンドグループあたりのダミーデータ数を変更することで、この発明を問題なく適用することができる。

【0047】

【発明の効果】この発明に依れば、MPEG圧縮された映像データとATRAC圧縮された音声データとを同期してメディアに記録することができる。

【0048】また、この発明に依れば、GOP単位で記録／再生を行うことができ、このときどのGOPから再生を開始しても映像と音声の同期（リップシンク）が守

られる。

【0049】さらに、この発明に依れば、1サウンドフレーム（約10.67ms）単位でアフレコを行うことができる。

【図面の簡単な説明】

【図1】この発明の一実施例のブロック図である。

【図2】この発明に係るMPEGで圧縮を行う映像データとATRACで圧縮を行う音声データとを説明するための略線図である。

*

*【図3】この発明に係るATRACエンコーダの一例を示す略線図である。

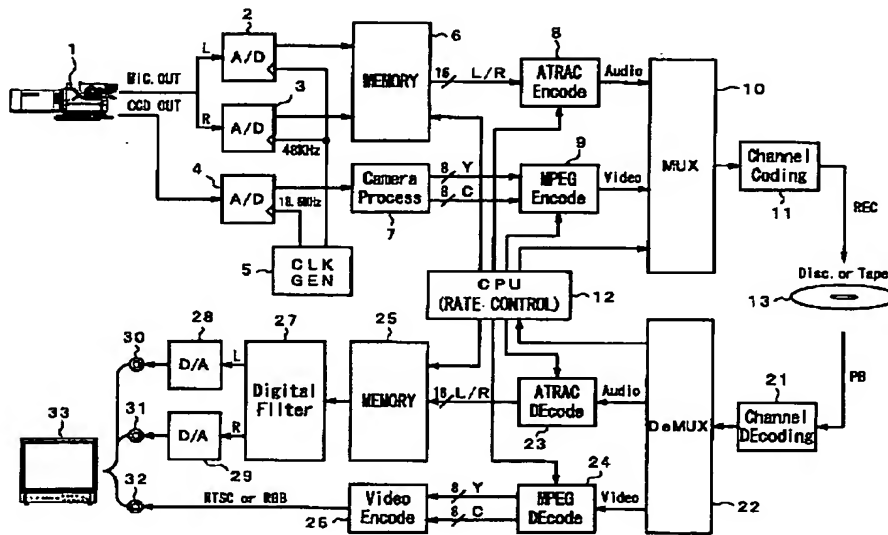
【図4】この発明に係るATRACエンコーダを説明するための略線図である。

【図5】この発明に係るATRACデコーダの一例を示す略線図である。

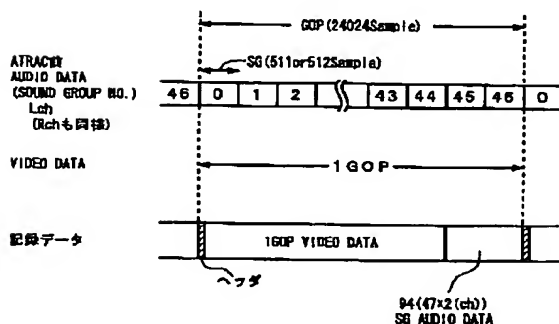
【符号の説明】

1・・・ビデオカメラ、2、3、4・・・A/D変換器、5・・・クロック発生器、6、25・・・メモリ、7・・・カメラプロセス回路、8・・・ATRACエンコーダ、9・・・MPEGエンコーダ、10・・・マルチプレクサ、11・・・チャンネルコーディング、12・・・CPU、13・・・記録媒体、21・・・チャンネルデコーディング、22・・・逆マルチプレクサ、23・・・ATRACデコーダ、24・・・MPEGデコーダ、26・・・ビデオエンコーダ、27・・・デジタルフィルタ、28、29・・・D/A変換器、33・・・TVモニタ

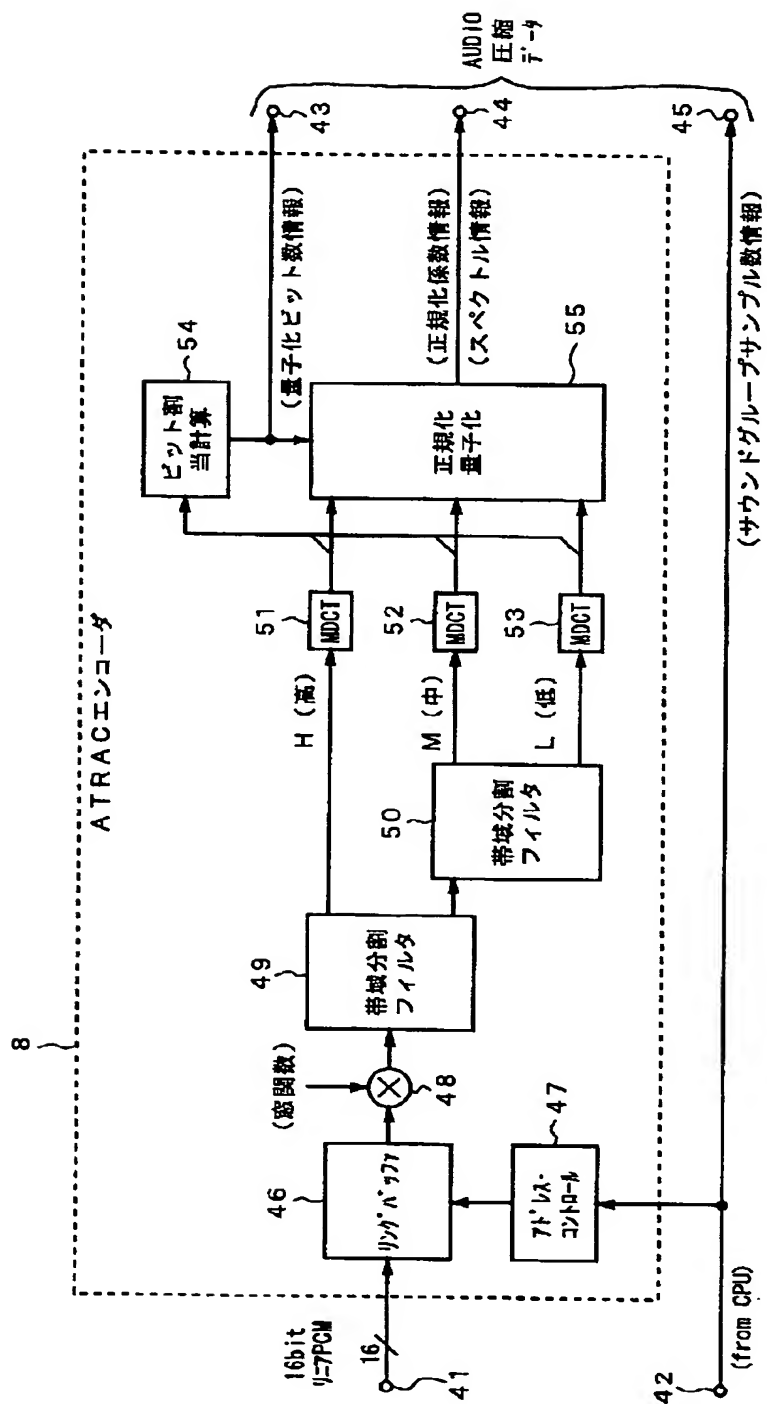
【図1】



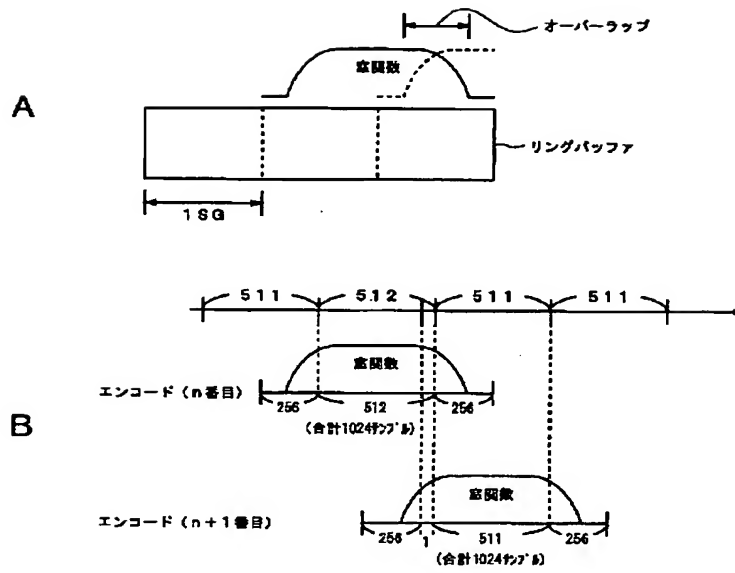
【図2】



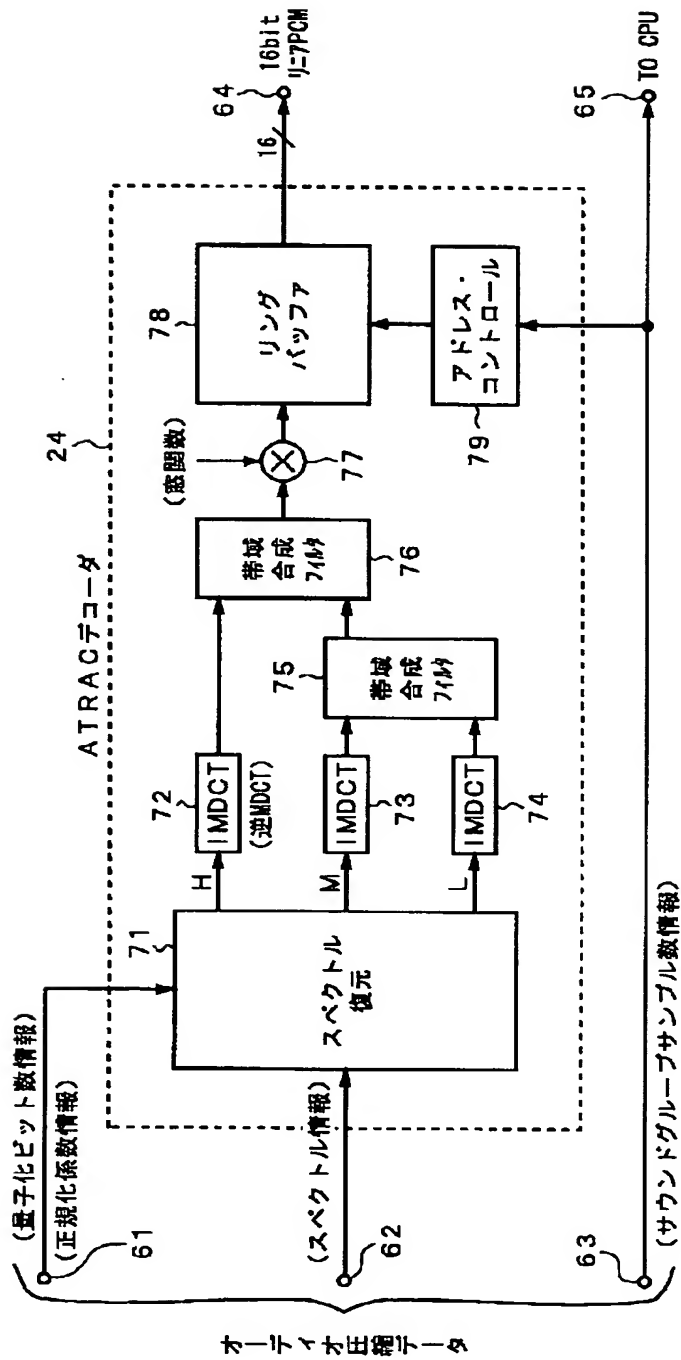
【図3】



【図4】



【図5】



フロントページの続き

(51)Int.Cl.⁶
H04N 7/24

識別記号

F I
H04N 7/13

Z